Journal of Statistical Mechanics:
Theory and Experiment

Derivation of Bose–Einstein statistics from the uncertainty principle

OPEN ACCESS

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Classical statistical mechanics, equilibrium and non-equilibrium

Derivation of Bose–Einstein statistics from the uncertainty principle

Paul Tangney

Department of Materials and Department of Physics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom
E-mail: p.tangney@imperial.ac.uk

Received 1 September 2023
Accepted for publication 20 August 2024
Published 24 September 2024

Online at stacks.iop.org/JSTAT/2024/093209
https://doi.org/10.1088/1742-5468/ad74e9

Abstract. The microstate of any degree of freedom of any classical dynamical system can be represented by a point in its two dimensional phase space. Since infinitely precise measurements are impossible, a measurement can, at best, constrain the location of this point to a region of phase space whose area is finite. This paper explores the implications of assuming that this finite area is bounded from below. I prove that if the same lower bound applied to every degree of freedom of a sufficiently-cold classical dynamical system, the distribution of the system's energy among its degrees of freedom would be a Bose–Einstein distribution.

Keywords: exact results, rigorous results in statistical mechanics, thermalization

Contents

1. Introduction

The development of quantum theory began with the discovery that energy radiating from a body at thermal equilibrium is not distributed among frequencies (f) as expected from (classical) statistical mechanics [1–3]. The only ways found to derive the experimentally-observed distribution involved assuming that either radiation itself, or the energy of an emitter of thermal radiation, was quantized into indivisible amounts hf, where $h \approx 6.6 \times 10^{-34}\ \mathrm{m^2 kg\,s^{-1}}$ became known as *Planck's constant* [1]. The distribution of energy among frequencies that this quantization implies became known as the *Bose–Einstein distribution*, in recognition of the refinement and extension of Planck's work by Bose and Einstein [4–6].

https://doi.org/10.1088/1742-5468/ad74e9

J. Stat. Mech. (2024) 093209

The discrepancy between the observed spectrum of a hot object and the expected one implied that the expectation was wrong. Planck's recognition that it could be resolved by assuming that light *emitters* have quantized energies [1] led Einstein to the conclusion that the energy of light itself is quantized [7]. Light quanta later became known as *photons* [8]. Here I show that the discrepancy can be resolved without concluding that either light itself, or emitters of light, have quantized energies. It can be resolved by assuming the existence of a universal lower bound on the precision to which the instantaneous microstate of any classically-evolving degree of freedom can be measured or known.

The Bose–Einstein distribution is generally regarded as among the most significant deviations of quantum physics from classical physics, and among the characteristics by which bosons differ from fermions. However the derivation of it presented here implies that any sufficiently-cold continuously-evolving classical dynamical system would be *observed* to obey Bose–Einstein statistics if the information provided by observations and measurements of it was limited by an uncertainty principle of the form $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_? > 0$, where Δ_{Q} and Δ_{P} are the uncertainties in the values of the canonically-conjugate variables that specify a microstate of a single degree of freedom.

1.1. Assumptions

Consider an arbitrary isolated continuously-evolving deterministic dynamical system, and let the instantaneous microstate of one of its degrees of freedom (DOF) be specified by a coordinate $Q_t \in \mathbb{Q}$ and the coordinate's conjugate momentum, $P_t \in \mathbb{P}$, where $\mathbb{Q} \cong \mathbb{R}$ and $\mathbb{P} \cong \mathbb{R}$ are the spaces of all possible coordinates and momenta, respectively.

Since the precision of any measurement is finite, the location of the point $\Gamma_t \equiv (Q_t, P_t)$ in the DOF's phase space, $\mathbb{G} \equiv \mathbb{Q} \times \mathbb{P} \cong \mathbb{R} \times \mathbb{R}$, cannot be known to infinite precision. There is always some degree of uncertainty in the values of Q_t and P_t. Therefore the only *certain* knowledge, as distinct from probabilistic knowledge, that an observer could possess about the point Γ_t is that it is *somewhere* in a specified finite-area subset of $\mathbb{G}$.

For simplicity, and because only the most accurate and precise measurements of the microstate are relevant to this work, let us assume that all measurements of Q_t and P_t result in the identifications of *interval* subsets of $\mathbb{Q}$ and $\mathbb{P}$, respectively, which are certain to contain them. Then all states of sufficiently-high certain knowledge about the location of Γ_t in $\mathbb{G}$ can be communicated as four values, Q, P, Δ_{Q}, and Δ_{P}, which specify a rectangular subset of $\mathbb{G}$ with vertices $(Q \pm \Delta_{\mathrm{Q}}/2, P \pm \Delta_{\mathrm{P}}/2)$ that is known to contain (Q_t, P_t).

The precisions, Δ_{Q} and Δ_{P}, to which Q_t and P_t can be determined depend in part on the microstate of the dynamical system at the time of measurement, and in part on how the measurement is performed.

J. Stat. Mech. (2024) 093209

1.1.1. Uncertainty principle. The unavoidably perturbative nature of the act of observation, and the fact that it is impossible for an observer to possess an infinite amount of information, imply that $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > 0$, but does not necessarily imply that the finite value of $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$ cannot be arbitrarily small.

However a finite Universe contains a finite amount of information. Therefore a lower bound on the value of $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$ must exist if the Universe is finite. For example, it is safe to say that the values of Δ_{Q} and Δ_{P} cannot be smaller than 10^{-N_p} in SI units, where N_p is the number of particles in the Universe. Therefore it is safe to say that there never has been, and never will be, a measurement of a DOF's microstate which determined the location of the microstate in its phase space to within an area of less than 10^{-2N_p} in SI units.

This extreme example demonstrates that $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$ would be bounded from below in a finite Universe, and its extremeness illustrates that larger universal lower bounds on microstate precision must also exist (e.g. $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > 10^{-N_p}$ in SI units). Only the largest of all universal lower bounds would be relevant to this work.

In an infinite Universe, it is not immediately obvious that $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$ cannot be arbitrarily small. However, uncertainty principles of the form $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_?$ arise in many contexts, and these uncertainty principles do not need to be universal (applicable in every possible context) for the derivation presented in this work to imply that, at low T, the distribution of a classical system's energy would appear to have the Bose–Einstein form in the context in which a particular uncertainty principle applies and is inviolable.

For example, [9] examines the relationship between macrostructure and microstructure, where *macrostructure* is the homogenized form that a microscopically-fluctuating classical field (the *microstructure*) is observed to have on a much larger time and/or length scale (the *macroscale*). It is shown that an uncertainty principle of the form $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_?$ applies at the macroscale when the probe used for all measurements is a macroscopic field (i.e. a homogenized microscopic field). Therefore if the direct or indirect source of all empirical knowledge was measurements with a macroscopic field, the energies of all sufficiently-cold classical systems would appear to be Bose–Einstein distributed.

Another example would be a dynamical system that was immersed in a bounded elastic medium. The wavelengths and frequencies (f) of classical waves in a bounded uniform medium are quantized, and the energy of a wave of amplitude A can be expressed as $\frac{1}{2}\gamma A^2 f^2$, for some medium-dependent constant γ. If Δf was the frequency quantum, the smallest energy difference between two waves, one of whose frequencies was f, would be

$$\Delta\mathscr{E} = \frac{1}{2}\gamma A^2 (f+\Delta f)^2 - \frac{1}{2}\gamma A^2 f^2 = \left(\gamma A^2 \Delta f\right) f + \mathcal{O}(\Delta f^2).$$

Therefore if all of an observer's knowledge about the immersed object had been communicated to them via the medium's waves, the smallest change in the energy of the object that could be communicated to them by observing the change in energy of a wave of frequency f would be $h_m f$, where $h_m = \gamma A^2 \Delta f$.

I base the otherwise-classical derivation presented in this work on the following nonstandard and strong assumption: There exists a finite lower bound $h_?$ on the value of $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$, and the same lower bound on microstate measurement precision applies to every observer and to every DOF of every classical dynamical system. In other words, I assume the existence of an uncertainty principle, $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_? > 0$, that is *universal*, meaning valid in every possible context. However, as discussed, if an uncertainty principle has a restricted validity, the derivation shares the uncertainty principle's domain of validity.

For the purposes of this work I will assume that all measurements of a DOF's microstate are performed at the lower bound on microstate precision, $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} = h_?$. Therefore each measurement of Γ_t reveals that it is in a rectangle of area $h_?$, centered at a point Γ, whose sides are parallel to the $\mathbb{Q}$ and $\mathbb{P}$ axes. I will denote such a rectangle by $\mathfrak{R}(\Gamma, \mathfrak{r})$, where $\mathfrak{r} \equiv \Delta_{\mathrm{Q}}/\Delta_{\mathrm{P}} = \Delta_{\mathrm{Q}}^2/h_? = h_?/\Delta_{\mathrm{P}}^2$.

I will use the assumption $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_? > 0$ to prove that, at thermal equilibrium, the distribution of any classical dynamical system's energy among its DOFs is a Bose–Einstein distribution in the low temperature (T) limit, albeit with $h_?$ in place of Planck's constant, h.

1.1.2. Low temperature limit. The low T limit is the limit in which Bose–Einstein statistics apply within quantum mechanics. Both classically and quantum mechanically, the low T limit is the weakly-interacting limit, and the Bose–Einstein distribution cannot be derived without assuming that interactions are weak enough to be approximated as absent for some purposes.

However it is important to clarify that assuming that an isolated physical system is in the low T limit means assuming that interactions are arbitrarily weak, but finite. It does not mean assuming that interactions are absent. This distinction is important because there would not be any energy exchange between DOFs if interactions were absent. Therefore a state of thermal equilibrium could never be reached, and it would not be meaningful to speak of the physical system having a temperature.

Let us assume that each DOF's pair of canonically-conjugate phase space coordinates, (Q_η, P_η), has been chosen such that, at any temperature, the Hamiltonian can be expressed exactly as

$$\mathcal{H}\left(\{(Q_\eta, P_\eta)\}\right) = U\left(\{Q_\eta\}\right) + \sum_\eta \mathcal{K}_\eta\left(P_\eta\right),$$

where U is the potential energy and $\mathcal{K}_\eta$ is the kinetic energy of DOF η.

Cooling the system brings its set of coordinates, $\{Q_\eta\}$, closer to a local minimum of U. Therefore, by reducing T, $\{Q_\eta\}$ can be brought arbitrarily close to a set, $\{Q_\eta^{\min}\}$, at which the partial derivative $\partial U/\partial Q_\eta$ vanishes for every η, and the second partial derivatives $\partial^2 U/\partial Q_\eta \partial Q_\mu$ are all either zero or positive. Furthermore, it is always possible to choose the set $\{Q_\eta\}$ such that the mixed derivatives, $\partial^2 U/\partial Q_\eta \partial Q_{\mu\neq\eta}$, vanish at $\{Q_\eta\} = \{Q_\eta^{\min}\}$. Therefore it is always possible to express the potential energy as

$$U = U^{\min} + \frac{1}{2}\sum_\eta \left.\frac{\partial^2 U}{\partial Q_\eta^2}\right|_{\{Q_\eta^{\min}\}} \Delta Q_\eta^2 + \mathcal{O}\left(\Delta Q^3\right),$$

J. Stat. Mech. (2024) 093209

where $U^{\min} \equiv U(\{Q_\eta^{\min}\})$ and $\Delta Q_\eta \equiv Q_\eta - Q_\eta^{\min}$.

Reducing T reduces the thermal averages of the ΔQ_η's and the standard deviation of their fluctuations. Therefore, by cooling to a sufficiently low T, the terms of orders ΔQ^3 and higher can be made negligible. This means that, by reducing T, the potential energy can be approximated arbitrarily closely as $U \approx U^{\min} + \sum_\eta U_\eta$, where $U_\eta = U_\eta(Q_\eta) \propto \Delta Q_\eta^2$. Therefore assuming that a physical system is in the low T limit allows the Hamiltonian to be approximated as

$$\mathcal{H} \approx U^{\min} + \sum_\eta \mathcal{H}_\eta(Q_\eta, P_\eta), \tag{1}$$

where $\mathcal{H}_\eta(Q_\eta, P_\eta) \equiv U_\eta(Q_\eta) + \mathcal{K}_\eta(P_\eta)$.

Since none of the terms on the right hand side of equation (1) depend on the phase space coordinates of more than one DOF, if the derivatives $\partial^2 U/\partial Q_\eta^2\big|_{\{Q_\eta^{\min}\}}$ are positive the DOFs only exchange energy through the neglected terms in the potential energy. These terms can be made arbitrarily small by reducing T, so interactions between DOFs can be made arbitrarily weak by reducing T.

If the derivatives $\partial^2 U/\partial Q_\eta^2$ are all zero, then each $\mathcal{H}_\eta$ is independent of Q_η, meaning that the system is gaseous. Therefore the DOFs only exchange energy during rare and brief 'collisions,' i.e. when the constant rates of change of two or more coordinates bring the set $\{Q_\eta\}$ into a region of the configuration space $\mathbb{Q}$ where U is not independent of the Q_η's. When that happens, the coordinates either condense into a set of weakly-interacting oscillators, or cease interacting again. If they cease interacting, their kinetic energies after the collision differ, in general, from their kinetic energies before the collision. If the duration of each collision is comparable to the time between collisions, T can be reduced until the former is a negligible fraction of the latter.

Regardless of whether a DOF becomes part of a set of weakly-interacting oscillators in the $T \to 0$ limit, or becomes an independent entity with constant potential energy in that limit, the assumption of a state of thermal equilibrium implies that energy *is* exchanged—either slowly or rarely. Therefore the equipartition theorem applies, which means that the time average of each DOF's energy is $\frac{1}{2}k_B T$, where k_B is Boltzmann's constant.

1.2. Outline of the derivation

The uncertainty principle is the only non-standard assumption that I make to show that the energy of every classical dynamical system is Bose–Einstein distributed in the $T \to 0$ limit. To derive this result, I take an information theoretical approach to statistical mechanics that is very similar to the one introduced, or championed, by Jaynes [10, 11]. Jaynes' approach leans heavily on the work of Shannon [12].

There are three important steps in the derivation. The first step, which I discuss in detail in section 2, is to recognize that, in the presence of uncertainty, the only empirically-unfalsifiable theories are statistical theories, and that the only empirically-unfalsifiable statistical theories are those in which uncertainty is maximised subject to

J. Stat. Mech. (2024) 093209

the constraint that everything that is known about the system is true. I refer to the set of all known information pertaining to a physical system as the system's *macrostate*.

The second step, which I discuss in detail in section 3, is to recognize that when an uncertainty principle applies, the domains of empirically testable probability distributions are quantized. The third step is to transform the coordinates (Q_η, P_η) canonically, such that $\mathcal{H}_\eta$ is transformed to a Hamiltonian with a particular form.

I will now outline the third step and explain why the derivation applies to *every* sufficiently cold classical dynamical system.

1.2.1. Transforming the Hamiltonian of each degree of freedom to an affine form. In section 4 I will show that when the uncertainty principle applies there is no inconsistency between the Bose–Einstein distribution and the Maxwell-Boltzmann distribution for a classical dynamical system: If the system is cold enough, the latter becomes the former under a canonical transformation of the set $\{(Q_\eta, P_\eta)\}$ of all phase space coordinates to a new set $\{(X_\eta, Y_\eta)\}$, which transforms the Hamiltonian of each DOF from $\mathcal{H}_\eta(Q_\eta, P_\eta) = U_\eta(Q_\eta) + \mathcal{K}_\eta(P_\eta)$ to one of the form

$$\tilde{\mathcal{H}} = \tilde{\mathcal{H}}_0 + \sum_\eta \tilde{\mathcal{H}}_\eta = \tilde{\mathcal{H}}_0 + \sum_\eta \left[B_\eta + C_\eta X_\eta\right],$$

where $\tilde{\mathcal{H}}_0$ is constant, and B_η, C_η, and Y_η are (approximately) constants of the motion of DOF η. Such a transformation is possible for *every* sufficiently-cold classical dynamical system at thermal equilibrium because, as discussed in section 1.1.2, each DOF is either a harmonic oscillator in that limit, or has a constant potential energy almost all of the time in that limit.

If the potential energy is constant, the only energy that the DOF can exchange with other DOFs is its kinetic energy, $\mathcal{H}_\eta = \mathcal{K}_\eta \propto P_\eta^2$. If Q_η oscillates harmonically, the energy of DOF η is proportional to the square of its oscillation amplitude, i.e. $\mathcal{H}_\eta \propto A_\eta^2$. Therefore $\mathcal{H}_\eta$ has the same mathematical form in each case, and this quadratic function of a single variable can be transformed canonically into an affine function of a single variable X_η whose form is $\tilde{\mathcal{H}}_\eta = B_\eta + C_\eta X_\eta$ [13].

For example, by transforming to *action-angle coordinates* $(Q_\eta, P_\eta) \mapsto (\mathcal{I}_\eta, \theta_\eta)$, the Hamiltonian of a set of harmonic oscillators is transformed from $\mathcal{H} = \frac{1}{2}\sum_\eta \left[P_\eta^2 + \omega_\eta^2 Q_\eta^2\right]$ to $\tilde{\mathcal{H}} = \sum_\eta \mathcal{I}_\eta \omega_\eta$, where $\omega_\eta = \dot{\theta}_\eta$ and the action $\mathcal{I}_\eta$ is a constant in the $T \to 0$ limit of arbitrarily weak interactions [14–16].

2. Unfalsifiable statistical models of deterministic systems

The purpose of this section is to explain the concept of an *unfalsifiable statistical model* of a classical Hamiltonian system. An example of such a model is the 19th century classical theory of thermodynamics. Some readers may wish to skip to section 4, and to return if or when they wish to scrutinise the logical foundations of the derivation more carefully.

I begin by explaining what I mean by an unfalsifiable statistical model. Then I explain my theoretical setup, before using this setup to derive the Maxwell-Boltzmann

J. Stat. Mech. (2024) 093209

distribution. In section 5 I show that, simply by changing the set of coordinates with which the microstate of a set of oscillators or waves is specified, the Maxwell-Boltzmann distribution becomes the Bose–Einstein distribution, albeit with an unknown constant in place of Planck's constant.

To understand what I mean by an unfalsifiable statistical theory or model, it is crucial to understand the difference between a *macrostate* and a *microstate*.

2.1. Macrostates and microstates

A classical *microstate* is complete information about the state of a deterministic system. It is a precise specification of the positions and momenta of all degrees of freedom of the system, or the values of any variables from which these positions and momenta could, in principle, be calculated.

A classical *microstructure* is complete information about the structure of a deterministic system, without any information about its rate of change with respect to time.

A *macrostate* $\mathcal{M}$ is simply a specification of the domain of applicability of a particular unfalsifiable statistical model. A macrostate is a set of information specifying everything that is known about the system to which the model applies. Because the model is statistical, it could only be falsified by a very large number of independent measurements. The macrostate is the complete list of everything that the samples on which these measurements are performed are known to have in common. It is also the complete list of everything that is known about each individual sample, and which may significantly influence the final reported result of the measurement, assuming that the uncertainty in this result is quantified correctly and reported with it.

2.2. Examples

2.2.1. Toy example. As a very simple example, let us suppose that $\mathcal{M}$ contains the following information only:

There are three lockable boxes, coloured red, green, and blue, at least one of which is unlocked. A ball has been placed inside one of the unlocked boxes. If more than one box is unlocked, the box into which the ball has been placed was chosen at random.

Let us suppose that an experiment on a system meeting specification $\mathcal{M}$ consists of an experimentalist checking which box the ball is in. Then, the only empirically-unfalsifiable statistical model of the experiment's results would be a probability distribution that assigns a probability of $\frac{1}{3}$ to the ball being in each box. Any other model could be falsified by statistics from an arbitrarily large number of repetitions of the experiment performed on independent realisations of system $\mathcal{M}$.

The fraction of times the ball would be found in each box would be $\frac{1}{3}$ even if different experiments were performed with different boxes locked, as long as the choice of which boxes were locked was made without bias, on average.

The model would be falsified by the empirical data if, say, the red box was chosen to be locked more frequently than the blue or green boxes. However, if that occurred, it would not mean that the unfalsifiable model was defective, but that it was being applied to the wrong macrostate. After the bias was discovered and quantified it would form part of the specification of a new macrostate, $\mathcal{M}'$, and an unfalsifiable statistical

J. Stat. Mech. (2024) 093209

theory of $\mathcal{M}'$ would be developed. Then, if no further macrostate-modifying peculiarities were found, the set of all subsequent repetitions of the experiment would produce data consistent with the unfalsifiable statistical theory of $\mathcal{M}'$.

2.2.2. Realistic example. While considering a more complicated example, it may be useful to have an infinite set of independent laboratories in mind. The equipment in each laboratory may be different, and different methods of measurement may be used in each one, but all are capable of measuring whatever quantities the unfalsifiable statistical model applies to. They are also capable of correcting their measurements for artefacts of the particular sample-preparation and measurement techniques they are using, and of accurately quantifying uncertainties in the corrected values.

Then one can imagine asking each laboratory to measure, say, the bulk modulus B of diamond at a pressure of 100 GPa and a temperature of 100 K. In this case, the statistical model would be a probability distribution, $p(B)$, for the bulk modulus of an infinitely large crystal (to eliminate surface effects, which are sample-specific) at precisely those values of pressure and temperature.

In general, each laboratory will prepare or acquire their sample of diamond in their own way, use a different method of controlling and measuring temperature and pressure, and use a different method of measuring B. In addition to the quantified uncertainties in the measured value of B, each independently-measured value will be influenced to some unquantified degree by *unknown unknowns*, i.e. unknown peculiarities of the sample, the apparatus, and the scientists performing the measurements and analysing the data. However, we will assume that this 'data jitter' either averages out, when the data from all laboratories is compiled, or is accounted for when comparing the compiled data to the statistical model.

If $p(B)$ was an unfalsifiable statistical model of B, it would be identical to the distribution of measured values. To derive or deduce an unfalsifiable distribution, one must carefully avoid making any assumptions, either explicitly or implicitly, about the sample or the measurement, apart from the information specified by the macrostate. This means maximising one's ignorance of every other property of a sample of diamond at $(P,T) = (100\,\mathrm{GPa}, 100\,\mathrm{K})$. This is achieved by maximising the uncertainty in the value of B that remains when its probability distribution, p, is known.

To derive an unfalsifiable distribution for a given macrostate, one must express the information specified by the macrostate as mathematical constraints on p. Then, under these information constraints, one must find the distribution p for which the uncertainty in the value of B is maximised. Maximising uncertainty eliminates bias and means that the information content of p is the same as the information content of the distribution of measured values of B. The differences between each distribution and a state of total ignorance is the same: it is the information about the value of B implied by the macrostate when no further information is available.

In summary, elimination of bias, subject to the constraint that information $\mathcal{M}$ is true, guarantees that the resulting statistical model of the physical system defined by $\mathcal{M}$ is unfalsifiable: It guarantees that the model would agree with a statistical model calculated from a very large amount of experimental data pertaining to physical systems about which $\mathcal{M}$, and only $\mathcal{M}$, is known to be true.

J. Stat. Mech. (2024) 093209

3. Probability domain quantization

The purpose of this section is to explain why one consequence of the uncertainty principle is that the most informative statistically unfalsifiable probability distribution for the location of a physical system's microstate in its phase space $\mathbb{G}$ is not a probability density function whose domain is $\mathbb{G}$, but a probability mass function whose domain is a partition of $\mathbb{G}$. In other words, the uncertainty principle quantizes the domain of any empirically testable probability distribution for the location of a classical dynamical system's microstate.

The derivations of the Maxwell-Boltzmann distribution in section 4 and the Bose–Einstein distribution in section 5 are reasonably self-contained, and reading this section is unnecessary to understand the gist of these derivations. However, skipping this section makes the derivations' logical foundations appear simpler than they are.

It also makes it *appear* that the derivations are built on an unjustified assumption: Namely, that an observer is capable of determining which point $\mathbf{\Gamma}$ on a lattice in phase space the microstate $\mathbf{\Gamma}_t$ of a physical system is closest to. This section will make clear that such an assumption is not made. It is important that it is not made because, as I now explain, the uncertainty principle, $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_?$, implies that an observer would not be capable of such a determination. Therefore that assumption would be a false premise.

If Γ and $\Gamma + \Delta\Gamma$ are adjacent points of the lattice in the phase space $\mathbb{G}$ of a single DOF, and if $\mathcal{N}_\Gamma$ and $\mathcal{N}_{\Gamma+\Delta\Gamma}$ are the sets of points in $\mathbb{G}$ that are closer to Γ and $\Gamma + \Delta\Gamma$, respectively, than to any other points of the lattice, then $\mathcal{N}_\Gamma$ and $\mathcal{N}_{\Gamma+\Delta\Gamma}$ share a border. The limit $h_?$ on microstate measurement precision implies that it is impossible for an observer to determine which side of their shared border the DOF's microstate Γ_t is on. Furthermore, as discussed in section 1.1, the result of a measurement of Γ_t is the identification of an element Γ of $\mathbb{G}$ and a ratio $\mathfrak{r} = \Delta_{\mathrm{Q}}/\Delta_{\mathrm{P}} \in \mathbb{R}^+$, such that $\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$. The value of Γ is not restricted to a point on a lattice, and the probability is zero that, by chance, it turns out to be one of the points of a particular lattice, because the measure of a lattice in $\mathbb{G}$ is zero. Therefore it is impossible for an observer to determine which point on a specific lattice Γ_t is closest to.

The purpose of section 3.1 is to discuss *probability spaces* that are capable of satisfying Kolmogorov's probability axioms [17–20]; and, in particular, some difficulties that arise when deriving a probability distribution for the location of Γ_t in $\mathbb{G}$. It is the uncertainty principle that causes the difficulties, and which forces us to confront certain subtleties in the definitions of probability spaces.

In order to resolve the difficulties, while ensuring empirical testability of the probability distributions that will be derived in sections 4 and 5, a detail will be added in section 3.2 to the infinite set of measurements (M measurements in the limit $M \to \infty$) performed on independently prepared physical systems that we imagined in section 2.

This detail is a filtration of the M results of those measurements: We will imagine defining an infinite set $\{p^{(C)}\}$ of different probability mass functions, $p^{(C)}$, each of which is consistent with a different subset of the $M \to \infty$ measurements, and each of whose domains is a different partition of $\mathbb{G}$. The introduction of this detail will clarify the true meaning of the apparently-false premise on which the derivations in sections 4 and 5 are built.

J. Stat. Mech. (2024) 093209

A brief clarification is that we can imagine that each function $p^{(\mathcal{C})}$ is assigned to a different agent, or 'statistician', and each probability mass function p that is derived in later sections is the function $p^{(\mathcal{C})}$ that has been assigned to one of those statisticians. This construction allows us to imagine calculating each distribution $p^{(\mathcal{C})}$ in two ways: The first is from the statistics gathered by the statistician to whom $p^{(\mathcal{C})}$ has been assigned ('Statistician $\mathcal{C}$'). The second is by using Jaynes' approach, as discussed in section 2, and as will be used in section 4, to theoretically derive the distribution that Statistician $\mathcal{C}$ would be unable to falsify.

3.1. Probability spaces

To develop a probabilistic description of the location of a microstate in its phase space, we must construct one or more probability spaces, each of which satisfy Kolmogorov's axioms of probability [17–20]. For simplicity, let us consider a physical system with a *single* DOF, whose microstate is $\Gamma_t \in \mathbb{G}$.

A probability space (S, Σ, P) consists of a *sample space* S, a *σ-algebra* Σ, and a *probability measure*, $P : \Sigma \to [0, 1]$. The sample space S is the set of all mutually-exclusive *outcomes* or results of a trial or measurement, and the σ-algebra Σ is the set of all events to which P assigns probabilities.

Σ is a *cover* of S, meaning that it is a collection of subsets whose union is the whole set. However, it is not necessary for the elements of S, which are the mutually exclusive outcomes, to be elements of Σ. The only properties of Σ that are required for the probability space to satisfy the axioms of probability are that it is a set of subsets of S, which includes S itself, and which is closed under countable unions ($A_1, A_2, \cdots \in \Sigma \implies \bigcup_{i=1}^{\infty} A_i \in \Sigma$), closed under countable intersections ($\bigcap_{i=1}^{\infty} A_i \in \Sigma$), and closed under complements ($A \in \Sigma \implies S \setminus A \in \Sigma$).

In section 3.1.1, we will consider the construction of a probability space for the outcome of a measurement of a microstate, in order to show that constructing it is straightforward.

In section 3.1.2 we will consider the construction of a probability space for the location of Γ_t in $\mathbb{G}$ in order to show that the uncertainty principle makes an unbiased construction of a single probability space impossible. To avoid introducing bias, it is necessary to introduce an infinite number of probability spaces.

In section 3.2, a logical construction will be outlined, which resolves some conceptual difficulties that arise when describing the location of Γ_t in $\mathbb{G}$ with an infinite number of probability distributions. This lays the logical foundations for the derivations of the Maxwell-Boltzmann and Bose–Einstein distributions presented in sections 4.2 and 5, respectively.

3.1.1. Probability space for the outcome of a measurement of Γ_t. The assumption made in section 1.1 was that the outcome of an accurate and maximally precise measurement of the location of Γ_t in $\mathbb{G}$ would be the identification of an element $(\Gamma, \mathfrak{r})$ of

$$\Omega \equiv \mathbb{G} \times \mathbb{R}^+ = \left\{ (\Gamma, \mathfrak{r}) : \Gamma \in \mathbb{G},\ \mathfrak{r} \in \mathbb{R}^+ \right\},$$

such that $\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$, and with no constraints placed on the values of $\Gamma \in \mathbb{G}$ and $\mathfrak{r} \in \mathbb{R}^+$ that might be discovered.

J. Stat. Mech. (2024) 093209

Let $\mathfrak{M}: S \to \Omega$ denote a random variable that maps the outcome o of a measurement of Γ_t to an element $\mathfrak{M}(o)$ of Ω; and let '$\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$' represent the outcome of the measurement that $\mathfrak{M}$ would map to the point $(\Gamma, \mathfrak{r}) \in \Omega$. Quotes are placed around $\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$ to indicate that '$\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$' represents the measurement outcome, which is a *revelation* and a *piece of information*, rather than the location of Γ_t that the information revealed implies.

The sample space for the measurement outcome is

$$\text{`}\Omega\text{'} \equiv \{\text{`}\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})\text{'} : (\Gamma, \mathfrak{r}) \in \Omega\}.$$

It is denoted by 'Ω' because its elements are in one-to-one correspondence with elements of Ω, and the quotes indicate that its elements are revelations, rather than locations.

The elements of 'Ω' are mutually exclusive because, assuming that $\Gamma_1 \neq \Gamma_2$ and/or $\mathfrak{r}_1 \neq \mathfrak{r}_2$, the outcome of the measurement can be '$\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1)$' or '$\Gamma_t \in \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$', but not both. It cannot be both because the result of each measurement is the revelation of a *single* imprecisely specified location. Since $\mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \neq \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$, '$\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1)$' and '$\Gamma_t \in \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$' are two different revelations, and both cannot occur.

The mutual exclusivity of elements of 'Ω' makes it straightforward to define a probability space $(\text{`}\Omega\text{'}, \wp(\text{`}\Omega\text{'}), P_o)$ for the measurement outcome, whose σ-algebra is the *power set* $\wp(\text{`}\Omega\text{'})$ of 'Ω'. A probability density function $\rho_o: \Omega \to \mathbb{R}^+$ can also be defined such that, for any $A \subseteq \Omega$,

$$P_o(\{o \in \text{`}\Omega\text{'} : \mathfrak{M}(o) \in A\}) = \int_A \mathrm{d}w \rho_o(w).$$

This is the probability that the measurement discovers that $\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$ for some $(\Gamma, \mathfrak{r}) \in A \subset \Omega = \mathbb{G} \times \mathbb{R}^+$. It is *not* the probability that Γ_t is in a particular subset of $\mathbb{G}$. However, in the limit $h_? \to 0^+$, we could define Ω as $\mathbb{G}$ instead of $\mathbb{G} \times \mathbb{R}^+$, in which case it would become such a probability.

Therefore, were it not for the uncertainty principle, it would not be necessary to draw attention to the distinction between the outcome $o \in$ 'Ω' of a measurement, and the element $\mathfrak{M}(o)$ of the measureable space Ω to which it is mapped by $\mathfrak{M}$. The uncertainty principle makes discussing this distinction important, because it is the location of Γ_t that we wish to model statistically, and we cannot directly use the range Ω of $\mathfrak{M}$ as the domain of a probability distribution for its location. We will explore the reasons for this next.

3.1.2. Probability space for the location of Γ_t in $\mathbb{G}$. To understand why it is not straightforward to define a probability space for the location of Γ_t, consider that, although the elements of 'Ω' are mutually exclusive, the elements of the set

$$\mathfrak{R}(\Omega) \equiv \{\mathfrak{R}(w) : w \in \Omega\},$$

which is the set of all imprecisely-specified locations of Γ_t that the measurement might discover, are not mutually exclusive locations. They are not mutually exclusive because $\mathfrak{R}(\Gamma_1, \mathfrak{r}_1)$ and $\mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ might intersect. If they intersected, $\mathfrak{R}(\Gamma_1, \mathfrak{r}_1)$ and $\mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ would not be mutually exclusive locations, but '$\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1)$' and '$\Gamma_t \in \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$' would still be mutually exclusive revelations because only one of them, at most, would be revealed.

J. Stat. Mech. (2024) 093209

The fact that the elements of $\mathfrak{R}(\Omega)$ are not mutually exclusive means that it cannot be treated as a sample space for the purpose of building a probability space. However, the problem is more serious than this: $\mathfrak{R}(\Omega)$ cannot even be a subset of a probability space's σ-algebra, because probabilities cannot be assigned to intersections of elements of $\mathfrak{R}(\Omega)$, and because a σ-algebra must be closed under intersections of countable numbers of its elements.

For example, a probability cannot be assigned to the event $\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$, despite the fact that there is an intuitively clear sense in which $\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ is a possibility. It cannot be assigned a probability because whether or not this possibility has been realised is unknowable. It is unknowable because the area of $\mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ is less than $h_?$, which means that to know that $\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ would imply a violation of the uncertainty principle.

Therefore, *in the context of defining the σ-algebra of a probability space*, the probability

$$\Pr\left(\Gamma_t \in \mathfrak{R}\left(\Gamma_1, \mathfrak{r}_1\right) \cap \mathfrak{R}\left(\Gamma_2, \mathfrak{r}_2\right)\right)$$

is not a meaningful quantity. The only probabilities that are meaningful in that context are the probabilities of events that can be known to have occurred or to have not occurred.

One illustration of the problems that would arise if $\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ was regarded as an event that could be assigned a finite probability is the fact that the probability measure that assigned the probability would be inconsistent with statistics gathered from an infinite number of measurements: The fraction of the measurements that would discover that event $\Gamma_t \in \mathfrak{R}(\Gamma_1, \mathfrak{r}_1) \cap \mathfrak{R}(\Gamma_2, \mathfrak{r}_2)$ had occurred would be zero.

Therefore a probability space whose probability measure would be consistent with an infinite number of measurements must be built from a sample space $\mathcal{C}$ that is a cover of $\mathbb{G}$ whose elements are mutually disjoint subsets of $\mathbb{G}$ of area no less than $h_?$. I use the term *disjoint* in the unconventional weaker sense that sets A and B are disjoint if the measure $|A \cap B|$ of their intersection is zero. Therefore elements of $\mathcal{C}$ may share boundaries.

Unfortunately, there are an infinite number of covers of $\mathbb{G}$ that meet these specifications. Therefore there are an infinite number of probability spaces that could be built for the location of Γ_t in $\mathbb{G}$, and choosing any one of them as the statistical model that describes Γ_t would be to introduce bias. For example, in general, the expectation value,

$$\langle O \rangle_{\mathcal{C}} \equiv \sum_{c \in \mathcal{C}} \Pr\left(\Gamma_t \in c\right) \left(\frac{1}{|c|} \int_c \mathrm{d}\Gamma O\left(\Gamma\right) \right),$$

depends on which cover $\mathcal{C}$ is chosen, where $|c|^{-1} \int_c \mathrm{d}\Gamma O(\Gamma)$ is the average on $c \subset \mathbb{G}$ of some function $O : \mathbb{G} \to \mathbb{R}$, and $|c|$ is the area of c.

To avoid bias, we must define, or be aware of the existence of, an infinite number of probability spaces: There is one probability space, $(\mathcal{C}, \wp(\mathcal{C}), P_{\mathcal{C}})$, and one probability distribution, $p^{(\mathcal{C})} : \mathcal{C} \to [0,1]; c \mapsto p^{(\mathcal{C})}(c) \equiv P_{\mathcal{C}}(c)$, for each cover $\mathcal{C}$.

The next step is to understand how each element of the infinite set $\{p^{(\mathcal{C})}\}$ of probability distributions could, in principle, be validated or invalidated by statistics from

an infinite set of measurements of Γ_t, each of whose outcomes is an element of 'Ω'. If it is not possible to imagine calculating a distribution from statistics, rather than deriving it theoretically, it cannot be claimed that the theoretically derived distribution is empirically unfalsifiable.

3.2. An infinitude of statisticians

Let us restrict attention to the most informative probability distributions possible. Therefore, let us disregard covers whose elements are larger than necessary, and only consider probability distributions $p^{(\mathcal{C})}$ whose domains are covers $\mathcal{C}$ whose elements all have areas of exactly $h_? + \delta h_?$. Let us also take the limit $\delta h_? \to 0^+$, so that $\delta h_?$ can be regarded as both finite and arbitrarily small. Let Λ denote the set of all covers that meet these specifications.

Now let us assume that the results of the $M \to \infty$ measurements are distributed among an infinite number of statisticians, such that there is exactly one statistician ('Statistician $\mathcal{C}$') for each $\mathcal{C} \in \Lambda$. Then let us imagine that each measurement whose outcome is '$\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$' is communicated to all of the statisticians whose covers contain an element of which $\mathfrak{R}(\Gamma, \mathfrak{r})$ is a subset, and is not communicated to the rest of the statisticians.

Clearly, every measurement of Γ_t determines that $\Gamma_t \in \mathbb{G}$. Therefore, by definition of a *cover*, every measurement determines that Γ_t is in *some* element of *every* statistician's cover. However, we are supposing that each statistician learns the result of each measurement of Γ_t if and only if the measurement has determined *which* element of their cover contains Γ_t. This can only be the case if the set $\mathfrak{R}(\Gamma, \mathfrak{r})$ that the measurement discovers Γ_t to be in is a subset of an element of their cover.

For each element c of $\mathcal{C}$, Statistician $\mathcal{C}$ calculates the fraction, $p^{(\mathcal{C})}(c)$, of the total number $M_{\mathcal{C}} < M$ of measurements whose outcomes they are privy to, for which $\Gamma_t \in c$. Therefore, in the limit $M \to \infty \implies M_{\mathcal{C}} \to \infty$, Statistician $\mathcal{C}$ calculates a probability distribution $p^{(\mathcal{C})}$, whose domain is $\mathcal{C}$.

The next question to address is the following: If one of the M measurements was chosen at random, is $p^{(\mathcal{C})}(c)$ the probability that c contains the microstate of the sample being measured in that measurement? In other words, is $\Pr(\Gamma_t \in c) = p^{(\mathcal{C})}(c)$?

The first thing to note is that $\Pr(\Gamma_t \in c)$ is an unknowable probability, for the same reason that, in general, $\Gamma_t \in c$ is an untestable proposition: $\mathcal{P}_{\text{certain}} \equiv P_o(\{\text{'}\Gamma_t \in c\text{'}\})$ is the fraction of the M measurements in which it is known that $\Gamma_t \in c$, and

$$\mathcal{P}_{\text{possible}} \equiv P_o\left(\{o \in \text{'}\Omega\text{'} : \mathfrak{R}(\mathfrak{M}(o)) \cap c \neq \emptyset\}\right)$$

is the fraction of the M measurements in which it is known that $\Gamma_t \in c$ is possible. However it is impossible to know the fraction of the M measurements for which $\Gamma_t \in c$, for reasons discussed in the introduction to section 3: the uncertainty principle implies that it can never be known which side of a border between elements of $\mathcal{C}$ Γ_t is on. Furthermore, because 'Ω' has an infinite number of elements, the ratio $\mathcal{P}_{\text{certain}}/\mathcal{P}_{\text{possible}}$ vanishes.

J. Stat. Mech. (2024) 093209

To shed more light on the empirically-unanswerable question of whether $\Pr(\Gamma_t \in c) = p^{(\mathcal{C})}(c)$, let us consider the possibility that

$$p^{(\mathcal{C}_1)}(c) \neq p^{(\mathcal{C}_2)}(c),$$

for two covers $\mathcal{C}_1, \mathcal{C}_2 \in \Lambda$ that both contain c. By construction, every time the measurement outcome is '$\Gamma_t \in c$', this outcome is revealed to both Statistician $\mathcal{C}_1$ and Statistician $\mathcal{C}_2$. Therefore the numbers of times that these statisticians learn that $\Gamma_t \in c$ are the same. Let us denote that number by m_c. Therefore

$$p^{(\mathcal{C}_1)}(c) \equiv \frac{m_c}{M_{\mathcal{C}_1}} \neq \frac{m_c}{M_{\mathcal{C}_2}} \equiv p^{(\mathcal{C}_2)}(c)$$

would imply that, even in the $M \to \infty$ limit, $M_{\mathcal{C}_1} \neq M_{\mathcal{C}_2}$. Therefore

$$M_{\mathcal{C}_1 \setminus \{c\}} = M_{\mathcal{C}_1} - m_c \neq M_{\mathcal{C}_2} - m_c = M_{\mathcal{C}_2 \setminus \{c\}},$$

where, for example, $M_{\mathcal{C}_1 \setminus \{c\}}$ is the number of times that it has been revealed to Statistician $\mathcal{C}_1$ that Γ_t is in an element of $\mathcal{C}_1$ that is not c. This implies that

$$\sum_{c' \in \mathcal{C}_1 \setminus \{c\}} p^{(\mathcal{C}_1)}(c') \neq \sum_{c' \in \mathcal{C}_2 \setminus \{c\}} p^{(\mathcal{C}_2)}(c').$$

Therefore, since $\mathcal{C}_1 \setminus \{c\}$ and $\mathcal{C}_2 \setminus \{c\}$ are both covers of $\mathbb{G} \setminus c$, $p^{(\mathcal{C}_1)}(c) \neq p^{(\mathcal{C}_2)}(c)$ would imply that the fraction of the M measurements that discover that Γ_t is in a particular subset of $\mathbb{G}$ of dimensions $\Delta_{\mathrm{Q}} \times \Delta_{\mathrm{P}} = h_? + \delta h_?$, would not be determined solely by the fraction of the M measured samples for which Γ_t is in that subset

In other words (and for clarity I will use the unjustifiable and unphysical assumption that it is possible to know the probabilities $\{\Pr(\Gamma_t \in g) : g \subset \mathbb{G}\}$), $p^{(\mathcal{C}_1)}(c) \neq p^{(\mathcal{C}_2)}(c)$ would imply that there does not exist a constant K such that $P_o(\text{'}\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})\text{'}) = K \Pr(\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r}))$ for all $(\Gamma, \mathfrak{r}) \in \Omega$.

Not only can we not rule out the possibility that K is not constant, it would be surprising if it were constant: It was mentioned in section 1.1 that the measurement precisions Δ_{Q} and Δ_{P} depend in part on Γ_t and in part on how the measurement of Γ_t is performed. Therefore, if the measurement outcome is '$\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$', the location of Γ_t in $\mathbb{G}$ has played a part in determining $\mathfrak{r}$, in general. The fact that it has also played a part in determining Γ is obvious.

However the dependence of K on the microstate of a physical system implies that K depends on the system's Hamiltonian, which implies that it depends on what the physical system is. In other words, this dependence cannot be a universal limitation on the act of measuring a DOF's microstate.

Therefore, instead of abandoning the prospect of devising a universally-applicable statistical model, such as Bose–Einstein statistics, this dependence should be treated as one of the pecularities of individual physical systems, or methods of measurement, that were discussed in section 2.2.2 and whose effects on statistics must be accounted for before those statistics can be compared with the predictions of universally-applicable statistical models. When deriving a statistical model that is universally applicable, it

J. Stat. Mech. (2024) 093209

is not only reasonable to assume that K is the same for every $(\Gamma, \mathfrak{r}) \in \Omega$, making that assumption appears to be unavoidable.

In other words, while bearing in mind that $p^{(\mathcal{C})}(c) \propto P_o(\text{'}\Gamma_t \in c\text{'}) \propto \Pr(\Gamma_t \in c)$ is an empirically untestable proposition, let us use it as a rough approximation to a more nuanced and precise interpretation of $p^{(\mathcal{C})}(c)$. Then, so that we can derive a universally-applicable statistical model, we purposely neglect pecularities of individual physical systems, and samples of those systems, because this is the only way to derive a model that is generally applicable. This entails assuming that the fraction of the $M_{\mathcal{C}}$ measurements revealed to Statistician $\mathcal{C}$ for which $\Gamma_t \in c$ equals the fraction of all M measurements for which $\Gamma_t \in c$.

3.2.1. Justification of a working assumption used in the derivations. As discussed above, $p^{(\mathcal{C})}(c) = \Pr(\Gamma_t \in c)$ is an empirically untestable proposition, but is also the only reasonable assumption to make when deriving a generally-applicable unfalsifiable probability distribution. It is equivalent to the assumption that the number of measurements whose outcome is '$\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$' is proportional to the number of measurements in which $\Gamma_t \in \mathfrak{R}(\Gamma, \mathfrak{r})$, with the same constant of proportionality for every Γ and every $\mathfrak{r}$.

Under the assumption that $p^{(\mathcal{C})}(c) = \Pr(\Gamma_t \in c)$, we can justify the working assumption that it is possible to determine which element of cover $\mathcal{C}$ contains Γ_t as follows: From the perspective of Statistician $\mathcal{C}$, the revelation of a measurement outcome to them can be regarded as their 'measurement' of Γ_t. Therefore, from their perspective, each of their measurements determines which element of $\mathcal{C}$ contains Γ_t.

Then we can imagine that Statistician $\mathcal{C}$ calculates $p^{(\mathcal{C})}(c)$ from the results of their 'measurements', and that if we are told the macrostate $\mathcal{M}$ that defines the measurements, we can theoretically derive a probability distribution whose domain is $\mathcal{C}$, and which agrees perfectly with $p^{(\mathcal{C})}(c)$, by eliminating all bias subject to the constraint that information $\mathcal{M}$ is true.

Each of the distributions derived in sections 4 and 5 can be interpreted as this theoretically-derived statistically-unfalsifiable probability distribution, where the statistics that fail to falsify it are those gathered by Statistician $\mathcal{C}$.

4. Derivation of an unfalsifiable energy distribution

Section 4.1 presents a brief summary of the theoretical setup that is used in sections 4.2 and 5 to derive the Maxwell-Boltzmann distribution and the Bose–Einstein distribution, respectively.

It is assumed that it is possible for a measurement to determine which element of a cover of a DOF's phase space, comprising disjoint subsets of area $h_?$, contains the DOF's microstate. Although this assumption is not compatible with the uncertainty principle discussed in section 1.1.1, its use in derivations as a *working assumption* was justified in section 3.2.

J. Stat. Mech. (2024) 093209

4.1. Theoretical setup

Consider an arbitrary continuously-evolving deterministic system whose microstate can be specified by $\mathbf{\Gamma} \equiv (\mathbf{P}, \mathbf{Q})$, where $\mathbf{Q} \equiv (Q_1, Q_2 \cdots)$ is some set of generalized coordinates and $\mathbf{P} \equiv (P_1, P_2 \cdots)$, where P_η is the momentum conjugate to Q_η. In this coordinate system, let $\mathcal{H}(\mathbf{\Gamma})$ denote the system's Hamiltonian, and, as before, $\mathbb{G} \equiv \mathbb{Q} \times \mathbb{P} \ni \mathbf{\Gamma}$, $\mathbb{Q} \ni \mathbf{Q}$, and $\mathbb{P} \ni \mathbf{P}$ denote the system's phase space, configuration space, and momentum space, respectively.

Let us begin by partitioning $\mathbb{G}$ into nonoverlapping subsets of equal measure (phase space 'volume') as follows: We choose a countable set $\mathcal{G}$ of evenly-spaced points (microstates) in $\mathbb{G}$ and define a neighbourhood $\mathcal{N}_{\mathbf{\Gamma}} \subset \mathbb{G}$ of each point $\mathbf{\Gamma} \in \mathcal{G}$ such that $\mathbb{G} = \bigcup_{\mathbf{\Gamma} \in \mathcal{G}} \mathcal{N}_{\mathbf{\Gamma}}$, and such that, if $\mathbf{\Gamma}, \mathbf{\Gamma}' \in \mathcal{G}$ are any two different points ($\mathbf{\Gamma} \neq \mathbf{\Gamma}'$), then $|\mathcal{N}_{\mathbf{\Gamma}} \cap \mathcal{N}_{\mathbf{\Gamma}'}| = 0$ and $|\mathcal{N}_{\mathbf{\Gamma}}| = |\mathcal{N}_{\mathbf{\Gamma}'}|$, where $|\mathcal{N}_{\mathbf{\Gamma}}|$ denotes the measure of $\mathcal{N}_{\mathbf{\Gamma}}$ in $\mathbb{G}$. For simplicity, let us assume that if $\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}}$, then $\mathbf{\Gamma}_t$ is closer to $\mathbf{\Gamma}$ than to any other element of $\mathcal{G}$. Therefore the interior of $\mathcal{N}_{\mathbf{\Gamma}}$ is the set of all points in $\mathbb{G}$ that are closer to $\mathbf{\Gamma}$ than to any other element of $\mathcal{G}$.

Now let $p_{\mathbf{\Gamma}}$, where $\mathbf{\Gamma} \in \mathcal{G}$, denote the probability, $\Pr(\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}})$, that $\mathbf{\Gamma}_t$ is within $\mathcal{N}_{\mathbf{\Gamma}}$. The probability distribution for the point $\mathbf{\Gamma}$ that identifies the region $\mathcal{N}_{\mathbf{\Gamma}}$ containing $\mathbf{\Gamma}_t$ is $p : \mathcal{G} \to [0,1]; \mathbf{\Gamma} \mapsto p_{\mathbf{\Gamma}}$.

Now let us suppose, momentarily, that $\mathbf{\Gamma}_t$ is known to be in region $\mathcal{N}_{\mathbf{\Gamma}}$, and that $\mathcal{N}_{\mathbf{\Gamma}}$ is partitioned into $W_{\mathbf{\Gamma}}$ nonoverlapping subsets of equal measure $v \equiv |\mathcal{N}_{\mathbf{\Gamma}}|/W_{\mathbf{\Gamma}}$. Then, as Shannon demonstrated [12], we can quantify the amount of information that must be revealed to determine which of these subsets $\mathbf{\Gamma}_t$ is in by $\log W_{\mathbf{\Gamma}} = \log |\mathcal{N}_{\mathbf{\Gamma}}| - \log v$. In the limit $W_{\mathbf{\Gamma}} \to \infty$, $v \to 0$, the quantity of information required becomes infinite. However, as discussed in section 1.1.1, we are assuming that v has a lower bound, which means that $W_{\mathbf{\Gamma}}$ has an upper bound.

Without losing generality, let us assume that these bounds are $|\mathcal{N}_{\mathbf{\Gamma}}|$ and 1, respectively. In other words, let us assume that when we originally partitioned $\mathbb{G}$, we chose the set $\mathcal{G}$ such that the following is true:

Given any microstate $\mathbf{\Gamma} \in \mathcal{G}$, and any microstate $\mathbf{\Gamma}' \in \mathbb{G}$, which is closer to $\mathbf{\Gamma}$ than to any other element of $\mathcal{G}$, it is theoretically possible to distinguish between $\mathbf{\Gamma}'$ and any element of $\mathcal{G} \setminus \{\mathbf{\Gamma}\}$ by empirical means; and it is impossible to distinguish between $\mathbf{\Gamma}'$ and $\mathbf{\Gamma}$ by empirical means.

I will refer to $\mathcal{G}$ as a *maximal set of mutually-distinguishable microstates*; I will refer to a sampling of $\mathbb{G}$ with such a set as a *maximal sampling*; and I will use $\mathbf{h} \equiv |\mathcal{N}_{\mathbf{\Gamma}}|$ to denote the measure of each neighbourhood $\mathcal{N}_{\mathbf{\Gamma}}$ in a maximal sampling of phase space.

4.2. Maxwell-Boltzmann statistics

This section draws heavily from the works of Jaynes [10] and Shannon [12].

Let us add the assumption that we know that the expectation value of the system's energy is $\mathscr{E}$. For example, the system might be a classical crystal whose average energy is determined by a heat bath to which it is coupled.

The system's state of thermal equilibrium can be defined as the probability distribution p that maximises the Shannon entropy [12], subject to the constraint that the Hamiltonian's expectation value,

$$\langle \mathcal{H} \rangle [p] \equiv \sum_{\mathbf{\Gamma} \in \mathcal{G}} p_{\mathbf{\Gamma}} \mathcal{H}(\mathbf{\Gamma}),$$

is equal to $\mathscr{E}$, and subject to the normalization constraint $\sum_{\mathbf{\Gamma} \in \mathcal{G}} p_{\mathbf{\Gamma}} = 1$. The Shannon entropy is

$$\langle S \rangle [p] \equiv \sum_{\mathbf{\Gamma} \in \mathcal{G}} p_{\mathbf{\Gamma}} \mathfrak{I}(p_{\mathbf{\Gamma}}), \tag{2}$$

where $\mathfrak{I}(p_{\mathbf{\Gamma}}) \equiv -\log p_{\mathbf{\Gamma}}$ is the Shannon information [12] of p at $\mathbf{\Gamma}$. From now on it will be implicit that $\sum_{\mathbf{\Gamma}}$ means $\sum_{\mathbf{\Gamma} \in \mathcal{G}}$.

The Shannon information, $\mathfrak{I}(p_{\mathbf{\Gamma}})$, quantifies how much would be learned, meaning by how much would the uncertainty in the location of $\mathbf{\Gamma}_t$ reduce, if it was discovered that $\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}}$. The functions $k\mathfrak{I}(p_{\mathbf{\Gamma}})$, for any $k \in \mathbb{R}^+$, are the only functions that satisfy the following three conditions: (i) they would vanish if it was known that $\mathbf{\Gamma}_t$ was in $\mathcal{N}_{\mathbf{\Gamma}}$ prior to 'discovering' it there, i.e. if $p_{\mathbf{\Gamma}} = 1$; (ii) they increase as the discovery that $\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}}$ becomes more surprising, i.e. as $p_{\mathbf{\Gamma}}$ decreases; and (iii) they are additive. Additivity means that if, for example, it was discovered that $\mathbf{\Gamma}_t$ was in either $\mathcal{N}_{\mathbf{\Gamma}}$ or $\mathcal{N}_{\mathbf{\Gamma}'}$ (i.e. $\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}} \cup \mathcal{N}_{\mathbf{\Gamma}'}$), the quantity of information about the location of $\mathbf{\Gamma}_t$ that was unknown would decrease by $\mathfrak{I}(p_{\mathbf{\Gamma}}) + \mathfrak{I}(p_{\mathbf{\Gamma}'})$.

Any probability distribution, p, is a state of knowledge that an observer could be in. The Shannon information, $\mathfrak{I}(p_{\mathbf{\Gamma}})$, of $p_{\mathbf{\Gamma}}$, quantifies the information that would be revealed by the discovery that $\mathbf{\Gamma}_t \in \mathcal{N}_{\mathbf{\Gamma}}$, and the Shannon entropy is the expectation value of the quantity of information that would be revealed by discovering which point $\mathbf{\Gamma}$ in the maximal set of mutually-distinguishable microstates $\mathcal{G}$ the true microstate $\mathbf{\Gamma}_t$ is closest to. Therefore $\langle S \rangle [p]$ quantifies the incompleteness of distribution p, as a state of knowledge, when the identity of the element of $\mathcal{G}$ that is closest to $\mathbf{\Gamma}_t$ is regarded as complete knowledge.

Whether or not $\langle S \rangle [p]$ is satisfactory as a quantification of uncertainty in all contexts is probably irrelevant in the present context, because we will be maximising its value subject to the stated contraints. Therefore what is relevant is that it increases monotonically as the location of $\mathbf{\Gamma}_t$ in $\mathbb{G}$ becomes more uncertain.

We can express the stationarity of $\langle S \rangle [p]$ subject to constraints $\langle \mathcal{H} \rangle [p] = \mathscr{E}$ and $\sum_{\mathbf{\Gamma}} p_{\mathbf{\Gamma}} = 1$ as

$$\delta \left\{ \langle S \rangle [p] - \beta \left(\langle \mathcal{H} \rangle [p] - \mathscr{E} \right) - \beta\lambda \left(\sum_{\mathbf{\Gamma}} p_{\mathbf{\Gamma}} - 1 \right) \right\} = 0,$$

where β and $\beta\lambda$ are Lagrange multipliers. If we divide across by $-\beta$ and define the constant $T \equiv (k_B \beta)^{-1}$, where k_B is the Boltzmann constant, this can be expressed as $\delta \left(\tilde{\mathcal{F}}[p] + \lambda \sum_{\mathbf{\Gamma}} p_{\mathbf{\Gamma}} \right) = 0$, where $\tilde{\mathcal{F}}[p] \equiv \langle \mathcal{H} \rangle [p] - k_B T \langle S \rangle [p]$. By taking a partial derivative

J. Stat. Mech. (2024) 093209

of $\tilde{\mathcal{F}}[p]+\lambda\sum_{\mathbf{\Gamma}}p_{\mathbf{\Gamma}}$ with respect to $p_{\mathbf{\Gamma}}$ and setting it equal to zero, we find that

$$p_{\mathbf{\Gamma}}=e^{-(\mathcal{H}(\mathbf{\Gamma})-\mathcal{F})/k_BT}=\mathcal{Z}^{-1}e^{-\mathcal{H}(\mathbf{\Gamma})/k_BT}, \tag{3}$$

where $\mathcal{Z}\equiv\exp\left(-\mathcal{F}/k_BT\right)$ is known as the *partition function* and we refer to the quantity $\mathcal{F}=-k_BT\log\mathcal{Z}$, which is the value taken by $\tilde{\mathcal{F}}[p]$ when it is stationary with respect to normalization-preserving variations of p, as the *free energy*.

Equation (3) is the familiar Maxwell-Boltzmann distribution and T is the temperature. The derivation of equation (3) is a derivation, based on the premises that precede it and those stated within it, of the only empirically-unfalsifiable probability distribution for the true microstate. It is unfalsifiable because it explicitly rejects bias by maximising uncertainty subject to one physical constraint, which is the only thing that we know about the state of the system; namely, that a heat bath ensures that its average energy is $\mathscr{E}$.

As discussed in section 2, the absence of bias guarantees us that if we had enough independent replicas of the physical system, and if the only thing we knew about each one was that its average energy was $\mathscr{E}$, and if we could determine by measurement which element $\mathcal{N}_{\mathbf{\Gamma}}$ of the phase space partition the microstate of each one was in, the fraction of those whose microstate was in $\mathcal{N}_{\mathbf{\Gamma}}$ would be $p_{\mathbf{\Gamma}}=e^{-\beta\mathcal{H}(\mathbf{\Gamma})}/\mathcal{Z}$.

Now let us make the simplifying assumption under which the Bose–Einstein distribution is valid within quantum mechanics: The total energy is a sum of the energies of *independent* DOFs. Within quantum mechanics these DOFs are often interpreted as particles.

With the Hamiltonian of DOF η denoted by $\mathcal{H}_\eta(\Gamma_\eta)$, where $\Gamma_\eta\equiv(Q_\eta,P_\eta)$, we can express the Hamiltonian of the set of all DOFs as

$$\mathcal{H}\left(\mathbf{\Gamma}\right)=\sum_{\eta}\mathcal{H}_\eta\left(\Gamma_\eta\right), \tag{4}$$

and we can express the partition function as

$$\mathcal{Z}\equiv\sum_{\mathbf{\Gamma}}e^{-\beta\mathcal{H}(\mathbf{\Gamma})}=\sum_{\mathbf{\Gamma}}\prod_{\eta}e^{-\beta\mathcal{H}_\eta(\Gamma_\eta)} \tag{5}$$

where the product $\prod_\eta$ is over all DOFs.

Now let us choose the maximal set of mutually-distinguishable microstates, $\mathcal{G}$, to be a lattice, which is the direct product $\prod_\eta^\times\mathcal{G}_\eta$, where $\mathcal{G}_\eta$ is both a two dimensional lattice and a maximal set of mutually-distinguishable points in the phase space $\mathbb{G}_\eta$ of DOF η. The area of the non-overlapping neighbourhoods $\mathcal{N}_{\Gamma_\eta}$ of Γ_η whose union is $\mathbb{G}_\eta$ is $h_?\equiv|\mathcal{N}_{\Gamma_\eta}|=\Delta Q_\eta\Delta P_\eta=\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}}$, where $\frac{1}{2}\Delta P_\eta$ is the smallest difference in momentum P_η between mutually-distinguishable microstates of η with the same coordinate; and $\frac{1}{2}\Delta Q_\eta$ is the smallest difference in coordinate Q_η between mutually-distinguishable microstates with the same momentum.

These choices and definitions allow us to swap the order of the sum and the product in equation (5), thereby expressing it as $\mathcal{Z}=\prod_\eta\mathcal{Z}_\eta$, where

$$\mathcal{Z}_\eta\equiv\sum_{\Gamma_\eta}e^{-\beta\mathcal{H}_\eta(\Gamma_\eta)}, \tag{6}$$

J. Stat. Mech. (2024) 093209

and where $\sum_{\Gamma_\eta}$ denotes $\sum_{\Gamma_\eta \in \mathcal{G}_\eta}$. If we know the partition function $\mathcal{Z}_\eta$ of each DOF η, we can calculate the partition function $\mathcal{Z}$ of the system as a whole.

In section 5 we will explore other ways to calculate $\mathcal{Z}$ by transforming away from $(\mathbf{P},\mathbf{Q})$ and (P_η,Q_η) to different sets of variables. To avoid a proliferation of new symbols, I will recycle the symbols $\mathbb{G}$, $\mathbb{G}_\eta$, $\mathcal{H}$, $\mathcal{H}_\eta$, $\mathbf{\Gamma}$, Γ_η, $\mathcal{N}_\mathbf{\Gamma}$, $\mathcal{G}$, $\mathcal{G}_\eta$, $\mathbf{h}$, $p_\mathbf{\Gamma}$, and p. They will have the same meanings in the new coordinates as they do for coordinates $(\mathbf{P},\mathbf{Q})$.

5. Bose–Einstein statistics

I will now derive the Bose–Einstein distribution for a classical system of non-interacting oscillators or standing waves. Then I will briefly discuss how the derivation can be generalized to other kinds of physical systems.

5.1. Oscillators and standing waves

As discussed in section 1.1.2, if the potential energy of a classical dynamical system is a smooth function $U(\mathbf{Q})$ of its microstructure $\mathbf{Q}$, the system can be brought arbitrarily close to a minimum of its potential energy, $\mathbf{Q}^{\min}$, by cooling it slowly. Once $||\mathbf{Q}-\mathbf{Q}^{\min}||$ is small enough, lowering its temperature further brings its dynamics closer to a superposition of harmonic oscillations of the normal modes of its stable equilibrium structure, $\mathbf{Q}^{\min}$. For example, a set of mutually-attractive particles would condense into a stable vibrating cluster when cooled. The normal modes of a finite crystal or a continuous bounded medium are standing waves, so their dynamics become superpositions of standing waves when they are cold enough.

If we specify the microstructure by the set of displacements from mechanical equilibrium along the normal mode eigenvectors, each DOF η is an oscillator or standing wave with a different angular frequency ω_η, in general, whose energy can be expressed as

$$\mathscr{E}_\eta \equiv \frac{1}{2}\left(\dot{Q}_\eta^2 + \omega_\eta^2 Q_\eta^2\right), \tag{7}$$

where the *mode coordinate* Q_η has the dimensions of distance $\times \sqrt{\text{mass}}$. In the limit $T \to 0$ the behaviour of the system is described perfectly by a Hamiltonian of the form

$$\mathcal{H}(\mathbf{Q},\mathbf{P}) = U\left(\mathbf{Q}^{\min}\right) + \frac{1}{2}\sum_\eta \left[P_\eta^2 + \omega_\eta^2 Q_\eta^2\right], \tag{8}$$

where $U(\mathbf{Q}^{\min})$ is a constant that is irrelevant to the dynamics, and $P_\eta \equiv \dot{Q}_\eta$ is the momentum conjugate to Q_η.

As illustrated in figure 1, the true path $\partial\mathcal{R}_\eta$ of mode η in its phase space $\mathbb{G}_\eta$ is continuous. It is only the accessible information about the path that is quantized. As discussed in section 4.2 and at the beginning of section 5, each point $\Gamma_\eta \in \partial\mathcal{R}_\eta$ is indistinguishable from all points within a neighbourhood of it, whose area is h_η.

Uncertainty manifests differently in the microstate probability distribution depending on which set of coordinates is used to specify the microstate. Having found that p is

J. Stat. Mech. (2024) 093209

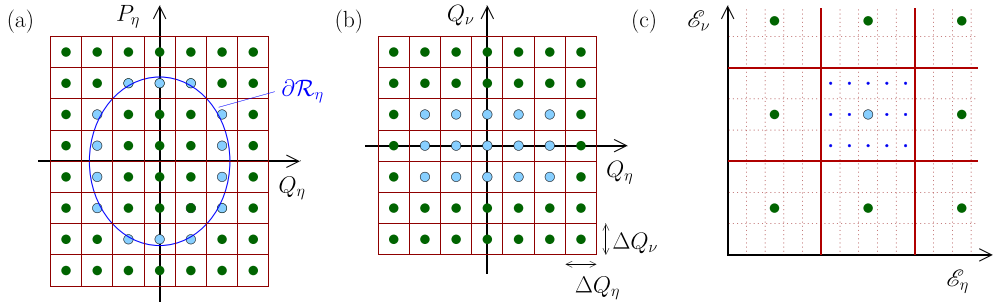

Figure 1. (a) A portion of the phase space $\mathbb{G}_\eta$ of mode η. The continuous blue ellipse, $\partial\mathcal{R}_\eta$, is a particular constant-energy path that the oscillation follows when it is decoupled from other modes. The set of pale blue and green spots is a maximal set of mutually-distinguishable microstates, Γ_η, of mode η. In statistical models of the mode's microstates, each spot represents all points within its rectangular neighbourhood. (b) A portion of the *microstructure* space of modes η and ν. The spots belong to a maximal set of mutually distinguishable points and represent the rectangular regions they inhabit. The pale blue spots mark regions visited during the motion of the modes, assuming that their energies, $\mathscr{E}_\eta$ and $\mathscr{E}_\nu$, are constant and that neither of their frequencies, ω_η and ω_ν, is an integer multiple of the other. Each of the 15 pale blue spots represents the four points $(Q_\eta, Q_\nu, P_\eta, P_\nu) = \left(Q_\eta, Q_\nu, \pm\sqrt{2\mathscr{E}_\eta - \omega_\eta^2 Q_\eta^2}, \pm\sqrt{2\mathscr{E}_\nu - \omega_\nu^2 Q_\nu^2}\right)$ in their joint phase space $\mathbb{G}_\eta \times \mathbb{G}_\nu$. (c) The pale blue spot is the energy of the trajectory represented by pale blue spots in panels (a) and (b). We cannot calculate the partition function of modes η and ν as $\mathcal{Z}_\eta\mathcal{Z}_\nu = \sum_{\mathscr{E}_\eta}\sum_{\mathscr{E}_\nu} e^{-\beta(\mathscr{E}_\eta+\mathscr{E}_\nu)}$ if the double summation is over a square grid in $(\mathscr{E}_\eta, \mathscr{E}_\nu)$-space. The numbers of energies sampled along each axis are only in the same ratio as the numbers of mutually-distinguishable mode coordinates along each axis in (Q_η, Q_ν)-space, and the numbers of mutually-distinguishable points in $\mathbb{G}_\eta$ and $\mathbb{G}_\nu$, if the spacings of sampled values along the mode's energy axes are their frequencies times the same constant.

a Maxwell-Boltzmann distribution when standard position and momentum coordinates (P_η, Q_η) are used, let us now perform the canonical transformation $(Q_\eta, P_\eta) \mapsto (\mathcal{I}_\eta, \vartheta_\eta)$, where $(\mathcal{I}_\eta, \vartheta_\eta)$ are the action-angle variables [14–16]. Then we will deduce the form of p when the microstate is specified as $\mathbf{\Gamma} = (\boldsymbol{\mathcal{I}}, \boldsymbol{\vartheta}) \equiv (\mathcal{I}_1, \mathcal{I}_2, \cdots, \vartheta_1, \vartheta_2, \cdots)$.

The action variable is defined as

$$\mathcal{I}_\eta \equiv \frac{1}{2\pi}\oint_{\partial\mathcal{R}_\eta} P_\eta \mathrm{d}Q_\eta = \frac{1}{2\pi}\int\int_{\mathcal{R}_\eta} \mathrm{d}P_\eta \wedge \mathrm{d}Q_\eta,$$

where the first integral is performed around the closed continuous trajectory $\partial\mathcal{R}_\eta$ defined by the equation $\mathcal{H}_\eta(Q_\eta, P_\eta) = \mathscr{E}_\eta$ and depicted in figure 1(a). The second expression, which involves an integral over the region $\mathcal{R}_\eta$ enclosed by the elliptical path $\partial\mathcal{R}_\eta$, follows from the generalized Stokes theorem.

It follows from the definition of $\mathcal{I}_\eta$ that $2\pi\mathcal{I}_\eta$ is the area enclosed by $\partial\mathcal{R}_\eta$. From equation (7), it is easy to see that the semi-axes of $\partial\mathcal{R}_\eta$ are $\sqrt{2\mathscr{E}_\eta}/\omega_\eta$ and $\sqrt{2\mathscr{E}_\eta}$.

J. Stat. Mech. (2024) 093209

Therefore, equating two expressions for the area enclosed gives

$$2\pi\mathcal{I}_\eta = \frac{2\pi}{\omega_\eta}\mathscr{E}_\eta \implies \mathcal{I}_\eta = \frac{\mathscr{E}_\eta}{\omega_\eta}.$$

The reason for choosing $\mathcal{I}_\eta$ as one of our variables should now be apparent: It allows us to express the new mode Hamiltonian as

$$\mathcal{H}_\eta(\vartheta_\eta, \mathcal{I}_\eta) = \mathcal{H}_\eta(\mathcal{I}_\eta) = \mathcal{I}_\eta\omega_\eta. \tag{9}$$

If we now followed precisely the same procedure with the new coordinates as we used to derive the Maxwell-Boltzmann distribution in section 4.2, we would reach equation (3), with $\mathcal{Z}$ and $\mathcal{Z}_\eta$ expressed as sums over all $\mathbf{\Gamma} \equiv (\boldsymbol{\mathcal{I}}, \boldsymbol{\vartheta}) \in \mathcal{G}$ and over all $\Gamma_\eta \in \mathcal{G}_\eta$, respectively. That is,

$$\mathcal{Z} = \sum_{\mathbf{\Gamma}} e^{-\beta\mathcal{H}(\mathbf{\Gamma})} = \sum_{\boldsymbol{\mathcal{I}}} e^{-\beta\mathcal{H}(\boldsymbol{\mathcal{I}})} = \prod_\eta \mathcal{Z}_\eta,$$

where

$$\mathcal{Z}_\eta \equiv \sum_{\mathcal{I}_\eta} e^{-\beta\mathcal{H}_\eta(\mathcal{I}_\eta)} = \sum_{\mathcal{I}_\eta} e^{-\beta\mathcal{I}_\eta\omega_\eta}, \tag{10}$$

and the sum over $\mathcal{I}_\eta$ is a sum over a maximal set, $\mathcal{G}_\eta \equiv \Delta\mathcal{I}_\eta\left(\mathbb{Z}_0^+ + \frac{1}{2}\right)$, of mutually-distinguishable values of $\mathcal{I}_\eta$. The reason for the factor $\frac{1}{2}$ is that the lower bound, $\frac{1}{2}\Delta\mathcal{I}_\eta$, on the difference between mutually-distinguishable values of $\mathcal{I}_\eta$ makes all points within the interval $[0, \frac{1}{2}\Delta\mathcal{I}_\eta)$ indistinguishable from zero, and makes zero indistinguishable from all points in this interval. Therefore, the sum in equation (10) can be viewed as $1/\Delta\mathcal{I}_\eta$ times a Riemann sum over $\mathbb{R}^+$, which samples intervals of width $\Delta\mathcal{I}_\eta$ centered at $\frac{1}{2}\Delta\mathcal{I}_\eta$, $\frac{3}{2}\Delta\mathcal{I}_\eta$, $\frac{5}{2}\Delta\mathcal{I}_\eta$, etc.

Now, since $h_?$ is a phase space area, the unavoidable uncertainty in the value of $2\pi\mathcal{I}_\eta$ must be $h_?$, and the unavoidable uncertainty in the value of $\mathcal{I}_\eta$ must be $\hbar_? \equiv h_?/(2\pi)$. Therefore the partition function can be expressed as

$$\begin{aligned}\mathcal{Z}_\eta &= \sum_{n_\eta \in \mathbb{Z}_0^+} e^{-\beta\left(n_\eta + \frac{1}{2}\right)\hbar_?\omega_\eta} \\ &= \frac{e^{-\frac{1}{2}\beta\hbar_?\omega_\eta}}{1 - e^{-\beta\hbar_?\omega_\eta}} = \frac{e^{\frac{1}{2}\beta\hbar_?\omega_\eta}}{e^{\beta\hbar_?\omega_\eta} - 1},\end{aligned}$$

where the second line has been reached by using the fact that the right hand side of the first line is an infinite geometric series. We can now express the free energy as

$$\begin{aligned}\mathcal{F} &= -\beta^{-1}\log\mathcal{Z} = -\beta^{-1}\sum_\eta \log\mathcal{Z}_\eta \\ &= \sum_\eta \left[\frac{1}{2}\hbar_?\omega_\eta + k_B T\log\left(1 - e^{-\beta\hbar_?\omega_\eta}\right)\right].\end{aligned}$$

J. Stat. Mech. (2024) 093209

The term $\frac{1}{2}\hbar_?\omega_\eta$ is commonly known as the *zero point energy* of mode η.

We can also calculate the expectation value,

$$\bar{n}_\eta \equiv \mathcal{Z}_\eta^{-1} \sum_{n_\eta \in \mathbb{Z}_0^+} n_\eta e^{-\beta\left(n_\eta+\frac{1}{2}\right)\hbar_?\omega_\eta}, \tag{11}$$

of n_η using equation (10) as follows:

$$\frac{\partial}{\partial\beta}\left(\sum_{n_\eta \in \mathbb{Z}_0^+} e^{-\beta\left(n_\eta+\frac{1}{2}\right)\hbar_?\omega_\eta}\right) = \frac{\partial}{\partial\beta}\left(\frac{e^{\frac{1}{2}\beta\hbar_?\omega_\eta}}{e^{\beta\hbar_?\omega_\eta}-1}\right).$$

After taking the derivatives and simplifying, this can be expressed as

$$\bar{n}_\eta = \frac{1}{e^{\beta\hbar_?\omega_\eta}-1}.$$

The integer n_η is commonly referred to as the *occupation number* of mode η and $\bar{n}_\eta$ is its thermal average.

When the modes' amplitudes are large enough that they do interact, their energies and frequencies vary, their paths in their phase spaces are no longer elliptical, and matters become more complicated. Nevertheless, simplifying assumptions are often justified, which allow a Bose–Einstein distribution to be used as the basis for a statistical description of the system's microstates and observables. For example, if the energy of mode η is modulated by a mode η' whose frequency is sufficiently low ($\omega_{\eta'} \ll \omega_\eta$), then $\mathcal{I}_\eta$ is approximately *adiabatically invariant* under this modulation [14–16], and the dominant effect of the interaction on mode η is to modulate its frequency.

As another example, when the interactions between modes are weak, the distribution of each mode's energy among frequencies is broadened and shifted relative to its $T \to 0$ limit. Therefore, it still has a well defined mean frequency and mean energy, which allows the Bose–Einstein distribution to be used effectively in many cases.

5.2. Generalizations to non-oscillatory systems

I have now derived the Bose–Einstein distribution for a classical system whose dynamics is a superposition of independent harmonic oscillations. My derivation made use of two properties of the system's Hamiltonian: The first was that it could be expressed as a sum $\mathcal{H} = \sum_\eta \mathcal{H}_\eta$ of the Hamiltonians $\mathcal{H}_\eta$ of independent DOFs. The second was that each $\mathcal{H}_\eta$ could be expressed as an affine function of only one variable. For oscillations, this was achieved by transforming to action-angle variables, so that each $\mathcal{H}_\eta$ took the form $\mathcal{H}_\eta(\mathcal{I}_\eta) = \mathcal{I}_\eta\omega_\eta$. Since variations of ω_η are negligible when interactions are weak, $\mathcal{I}_\eta$ is effectively the only variable that appears in $\mathcal{H}_\eta$.

The Hamiltonians of many other kinds of physical systems, composed of mutually-noninteracting DOFs, can be transformed canonically into forms that allow the Bose–Einstein distribution to be derived. In principle it can be derived whenever there exists a curve $\gamma_\eta : \mathbb{R}^+ \to \mathbb{G}_\eta; t \mapsto \gamma_\eta(t)$ in the phase space $\mathbb{G}_\eta$ of each DOF such that energies of DOF η are represented on γ_η in the same proportions as they are represented in $\mathbb{G}_\eta$.

J. Stat. Mech. (2024) 093209

To be more precise, energies should be represented on *maximal samplings* of γ_η in the same proportions as they are represented on maximal samplings of $\mathbb{G}_\eta$.

Once each $\mathcal{H}_\eta$ has been transformed canonically into the form $\mathcal{H}_\eta(X_\eta) = B_\eta + C_\eta X_\eta$, where $X_\eta \in \mathbb{R}$ is a continuously-varying generalized coordinate or momentum, and B_η and C_η are constants, the full Hamiltonian becomes

$$\mathcal{H}(\mathbf{X}) \equiv U\left(\mathbf{Q}^{\min}\right) + \sum_\eta B_\eta + \sum_\eta C_\eta X_\eta \tag{12}$$

where $\mathbf{X} \equiv (X_1, X_2, \cdots)$. Let $\frac{1}{2}\Delta X_\eta$ denote the smallest difference between mutually-distinguishable values of X_η; let $\epsilon_\eta \equiv C_\eta \Delta X_\eta$; and let $D \equiv e^{-\beta\left(U(\mathbf{Q}^{\min}) + \sum_\eta B_\eta\right)}$. Then the partition function can be expressed as

$$\mathcal{Z} = D \prod_\eta \sum_{n_\eta \in \mathbb{Z}_0^+} e^{-\beta\left(n_\eta + \frac{1}{2}\right)\epsilon_\eta},$$

and it is straightforward to show that $\bar{n}_\eta = 1/\left(e^{\beta\epsilon_\eta} - 1\right)$.

One example of a system whose Hamiltonian can be transformed canonically into the form of equation (12) is an ideal gas. At any given point in time, its Hamiltonian has the form, $\mathcal{H}(\mathbf{P}) \equiv \sum_\eta \mathcal{H}_\eta(P_\eta) = \frac{1}{2}\sum_\eta P_\eta^2$, which is the Hamiltonian of a set of independent free particles. A free particle Hamiltonian can be transformed canonically into a harmonic oscillator Hamiltonian [13]; therefore, it can also be transformed into action-angle coordinates.

As discussed in section 1.1.2, in the limit $T \to 0$ the Hamiltonian of *every* classical dynamical system either takes the same form as a set of weakly interacting harmonic oscillators or as an ideal gas, or as a combination of both. Therefore, all classical dynamical systems that are subject to an uncertainty principle, $\Delta_{\mathrm{Q}}\Delta_{\mathrm{P}} > h_? > 0$, obey Bose–Einstein statistics in the $T \to 0$ limit as a consequence of the probability domain quantization discussed in section 3.

As the derivation of the Bose–Einstein distribution presented in section 5.1 makes clear, when an uncertainty principle applies, the Maxwell-Boltzmann and Bose–Einstein distributions are perfectly compatible with one another in the $T \to 0$ limit. In that limit, a classical system's energy distribution can be expressed either as a Maxwell-Boltzmann distribution or as a Bose–Einstein distribution, depending on which choice of coordinates and their conjugate momenta are used to specify the microstate.

6. Discussion

I have shown that the Bose–Einstein distribution follows mathematically from *probability domain quantization*, and that probability domain quantization is a consequence of the existence of a limit, $h_?$, on the precision with which a system's microstate can be determined experimentally. Probability domain quantization does not imply that the microstates of the underlying physical system are quantized. It implies a quantization of the information contained in probability distributions that possess the quality of being testable empirically.

J. Stat. Mech. (2024) 093209

I have not justified my working assumption that a lower bound $h_?$ exists, but only demonstrated that one of its consequences would be that all sufficiently-cold classical dynamical systems are described by Bose–Einstein statistics. Therefore I have demonstrated that the existence of such a lower bound would have many important implications.

One implication would be that there is no qualitative discrepancy between the experimentally-observed spectrum of a blackbody and what should be expected if light was a mechanical wave in a bounded medium. As discussed in section 1.1.1, if light was such a wave, the boundedness of the medium would mean that the smallest energy difference between two light waves of frequency $\approx f$ would be $h_m f$, for some constant h_m. In section 5 I showed that the lower bound that an uncertainty principle would place on *observable* energy differences would be $\Delta(\mathcal{I}\omega) = (\Delta\mathcal{I})\,\omega = h_? f$. Therefore, if $h_m = h_?$, the existence of an uncertainty principle in a classical Universe could be explained by all observations being mediated by classical light waves.

Another implication of a lower bound $h_?$ would be that there is no qualitative discrepancy between the experimentally-observed temperature dependence of a crystal's heat capacity and what should be expected of classical lattice waves.

Another implication would be that classical oscillators and waves would have *zero point energies* that were simply an artefact of small energies being empirically-indistinguishable from no energy.

Another implication would be that, when a cluster of massive particles was cold enough, the classical expectation would be that almost all of its vibrational energy would be possessed by its lowest-frequency normal mode. Therefore, below a certain temperature, all but one of its degrees of freedom would be almost inactive and it would be a Bose–Einstein condensate.

For simplicity I have assumed that the limit on microstate measurement precision that leads to probability domain quantization is a limit on *certain* knowledge. In other words, I assumed that it is theoretically possible to know *with certainty* that a DOF's microstate is within a subset of its phase space if and only if the area of that subset is greater than $h_?$. If, instead, it is assumed that the result of the most precise microstate measurements possible are probability density functions of the form

$$\rho(\sigma_{\mathrm{Q}}, \sigma_{\mathrm{P}}) : \mathbb{Q} \times \mathbb{P} \to \mathbb{R}^{+};\ (Q, P) \mapsto \rho(Q, P; \sigma_{\mathrm{Q}}, \sigma_{\mathrm{P}}),$$

where σ_{Q} and σ_{P} are the standard deviations along the coordinate axis $\mathbb{Q}$ and the momentum axis $\mathbb{P}$, respectively, a more general form of uncertainty principle would be $\sigma_{\mathrm{Q}}\sigma_{\mathrm{P}} > h_?$. This would be a limit on *probabilistic* knowledge. It may be possible to adapt the derivations presented in this work to uncertainty principles of this more general form.

References

[1] Planck M 1901 Ueber das gesetz der energieverteilung im normalspectrum *Ann. Phys., Lpz.* **309** 553
[2] Mehra J and Rechenberg H 2001 *The Historical Development of Quantum Theory* vol I (Springer)
[3] Gorroochurn P 2018 The end of statistical independence: the story of Bose-Einstein statistics *Math. Intell.* **40** 12
[4] Bose B. 1924 Plancks gesetz und lichtquantenhypothese *Z. Phys.* **26** 178

J. Stat. Mech. (2024) 093209

[5] Einstein A 1924 Sitzungsberichte der preussischen akademie der wissenschaften *Math. Phys. Klass.* **261** 1925
[6] Einstein A 1925 Sitzungsberichte der preussischen akademie der wissenschaften *Math. Phys. Klass.* **3**
[7] Einstein A 1905 Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt *Ann. Phys.* **322** 132
[8] Lewis G N 1926 The conservation of photons *Nature* **118** 874
[9] Tangney P 2024 Electricity at the macroscale and its microscopic origins (arXiv:2403.13981)
[10] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
[11] Jaynes E T 1957 Information theory and statistical mechanics.II *Phys. Rev.* **108** 171
[12] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379
[13] Glass E N and Scanio Joseph J. G. 1977 Canonical transformation to the free particle *Am. J. Phys.* **45** 344
[14] Landau L and Lifshitz E 1976 *Mechanics (Course of Theoretical Physics* vol 1 Elsevier Science)
[15] Arnold V I 1989 *Mathematical Methods of Classical Mechanics* (*Graduate Texts in Mathematics*) 2nd edn (Springer)
[16] Lanczos C 1949 *The Variational Principles of Mechanics* (*Heritage Series*) (University of Toronto Press)
[17] Kolmogorov A N 1960 *Foundations of the Theory of Probability* 2nd edn (Chelsea Pub Co)
[18] Borovkov A A 2013 *Probability Theory* (*Universitext* (Springer))
[19] Varadhan S R S 2001 *Probability Theory (Courant Lecture Notes)* (Courant Institute of Mathematical Sciences)
[20] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge University Press)

J. Stat. Mech. (2024) 093209